Lecture 7: October 18, 2016

Lecturer: Madhur Tulsiani

1 Singular Value Decomposition

Let V, W be finite-dimensional inner product spaces and let $\varphi : V \to W$ be a linear transformation. Since the domain and range of φ are different, we cannot analyze it in terms of eigenvectors. However, we can use the spectral theorem to analyze the operators $\varphi^*\varphi : V \to V$ and $\varphi\varphi^* : W \to W$ and use their eigenvectors to derive a nice decomposition of φ. This is known as the singular value decomposition (SVD) of φ.

Proposition 1.1 *Let $\varphi : V \to W$ be a linear transformation. Then $\varphi^*\varphi : V \to V$ and $\varphi\varphi^* : W \to W$ are positive semidefinite linear operators with the same non-zero eigenvalues.*

In fact, we can notice the following from the proof of the above proposition.

Proposition 1.2 *Let v be an eigenvector of $\varphi^*\varphi$ with eigenvalue $\lambda \neq 0$. Then $\varphi(v)$ is an eigenvector of $\varphi\varphi^*$ with eigenvalue λ. Similarly, if w is an eigenvector of $\varphi\varphi^*$ with eigenvalue $\lambda \neq 0$, then $\varphi^*(w)$ is an eigenvector of $\varphi^*\varphi$ with eigenvalue λ.*

Using the above, we get the following.

Proposition 1.3 *Let $\sigma_1^2 \geq \sigma_2^2 \geq \cdots \geq \sigma_r^2 > 0$ be the non-zero eigenvalues of $\varphi^*\varphi$, and let $v_1, \ldots, v_r$ be a corresponding orthonormal eigenbasis. For $w_1, \ldots, w_r$ defined as $w_i = \varphi(v_i)/\sigma_i$, we have that*

1. *$\{w_1, \ldots, w_r\}$ form an orthonormal set.*
2. *For all $i \in [r]$*
$$\varphi(v_i) = \sigma_i \cdot w_i \quad \text{and} \quad \varphi^*(w_i) = \sigma_i \cdot v_i .$$

The values $\sigma_1, \ldots, \sigma_r$ are known as the (non-zero) singular values of φ. For each $i \in [r]$, the vector v_i is known as the right singular vector and w_i is known as the left singular vector corresponding to the singular value σ_i.

Proposition 1.4 *Let r be the number of non-zero eigenvalues of $\varphi^* \varphi$. Then,*

$$\mathrm{rank}(\varphi) \;=\; \dim(\mathrm{im}(\varphi)) \;=\; r\,.$$

Using the above, we can write φ in a particularly convenient form. We first need the following definition.

Definition 1.5 *Let V, W be inner product spaces and let $v \in V, w \in W$ be any two vectors. The* outer product *of w with v, denoted as $|w\rangle\langle v|$, is a linear transformation from V to W such that*

$$|w\rangle\langle v|\,(u) \;:=\; \langle u, v\rangle \cdot w\,.$$

Note that if $\|v\| = 1$, then $|w\rangle\langle v|\,(v) = w$ and $|w\rangle\langle v|\,(u) = 0$ for all $u \perp v$. Also, note that the rank of the linear transformation defined above is 1. We can then write $\varphi : V \to W$ in terms of outer products of its singular vectors.

Proposition 1.6 *Let V, W be finite dimensional inner product spaces and let $\varphi : V \to W$ be a linear transformation with non-zero singular values $\sigma_1, \ldots, \sigma_r$, right singular vectors $v_1, \ldots, v_r$ and left singular vectors $w_1, \ldots, w_r$. Then,*

$$\varphi \;=\; \sum_{i=1}^{r} \sigma_i \cdot |w_i\rangle\langle v_i|\;.$$

2 Singular Value Decomposition for matrices

Using the previous discussion, we can write matrices in convenient form. Let $A \in \mathbb{C}^{m \times n}$, which can be thought of as an operator from $\mathbb{C}^n$ to $\mathbb{C}^m$. Let $\sigma_1, \ldots, \sigma_r$ be the non-zero singular values and let $v_1, \ldots, v_r$ and $w_1, \ldots, w_r$ be the right and left singular vectors respectively. Note that $V = \mathbb{C}^n$ and $W = \mathbb{C}^m$ and $v \in V, w \in W$, we can write the operator $|w\rangle\langle v|$ as the matrix wv^*, there v^* denotes $\overline{v^T}$. This is because for any $u \in V$, $wv^*u = w(v^*u) = \langle u, v\rangle \cdot w$. Thus, we can write

$$A \;=\; \sum_{i=1}^{r} \sigma_i \cdot w_i v_i^*\,.$$

Let $W \in \mathbb{C}^{m \times r}$ be a matrix with $w_1, \ldots, w_r$ as columns, such that i^{th} column equals w_i. Similarly, let $V \in \mathbb{C}^{n \times r}$ be a matrix with $v_1, \ldots, v_r$ as the columns. Let $\Sigma \in \mathbb{C}^{r \times r}$ be a diagonal matrix with $\Sigma_{ii} = \sigma_i$. Then, check that the above expression for A can also be written as

$$A \;=\; W \Sigma V^*\,,$$

where $V^* = \overline{V^T}$ as before.

We can also complete the bases $\{v_1, \ldots, v_r\}$ and $\{w_1, \ldots, w_r\}$ to bases for $\mathbb{C}^n$ and $\mathbb{C}^m$ respectively and write the above in terms of unitary matrices.

Definition 2.1 *A matrix $U \in \mathbb{C}^{n\times n}$ is known as a* unitary matrix *if the columns of U form an orthonormal basis for $\mathbb{C}^n$.*

Proposition 2.2 *Let $U \in \mathbb{C}^{n\times n}$ be a unitary matrix. Then $UU^* = U^*U = \mathsf{id}$, where id denotes the identity matrix.*

Let $\{v_1, \ldots, v_n\}$ be a completion of $\{v_1, \ldots, v_r\}$ to an orthonormal basis of $\mathbb{C}^n$, and let $V_n \in \mathbb{C}^{n\times n}$ be a unitary matrix with $\{v_1, \ldots, v_n\}$ as columns. Similarly, let $W_m \in \mathbb{C}^{m\times m}$ be a unitary matrix with a completion of $\{w_1, \ldots, w_r\}$ as columns. Let $\Sigma' \in \mathbb{C}^{m\times n}$ be a matrix with $\Sigma'_{ii} = \sigma_i$ if $i \leq r$, and all other entries equal to zero. Then, we can also write

$$A = W_m \Sigma' V_n^* .$$

2.1 SVD as a low-rank approximation for matrices

Given a matrix $A \in \mathbb{C}^{m\times n}$, we want to find a matrix B of rank at most k which "approximates" A. For now we will consider the notion of approximation in spectral norm i.e., we want to minimixe $\|A - B\|_2$, where

$$\|(A-B)\|_2 \;=\; \inf_{x\neq 0} \frac{\|(A-B)x\|_2}{\|x\|_2} .$$

SVD also gives the optimal solution for another notion of approximation: minimizing the Frobenius norm $\|A - B\|_F$, which equals $(\sum_{ij}(A_{ij} - B_{ij})^2)^{1/2}$. We will see this later. Let $A = \sum_{i=1}^{r} w_i v_i^*$ be the singular value decomposition of A and let $\sigma_1 \geq \cdots \geq \sigma_r > 0$. If $k \geq r$, we can simply use $B = A$ since $\mathrm{rank}(A) = r$. If $k < r$, we claim that $A_k = \sum_{i=1}^{k} \sigma_i w_i v_i^*$ is the optimal solution. If is easy to check the following.

Proposition 2.3 $\|A - A_k\|_2 \;=\; \sigma_{k+1}$.

Thus, we know that the error of the best approximation B is at most σ_{k+1}. To show the lower bound, we need the following fact.

Exercise 2.4 *Let V be a finite-dimensional vector space and let S_1, S_2 be subspaces of V. Then, $S_1 \cap S_2$ is also a subspace and satisfies*

$$\dim(S_1 \cap S_2) \;\geq\; \dim(S_1) + \dim(S_2) - \dim(V) .$$

We can now show the following.

Proposition 2.5 *Let $B \in \mathbb{C}^{m\times n}$ have* $\mathrm{rank}(B) \leq k$ *and let $k < r$. Then $\|A - B\|_2 \geq \sigma_{k+1}$.*

Proof: By rank-nullity theorem $\dim(\ker(A)) \geq n-k$. Thus, by the fact above

$$\dim\left(\ker(A) \cap \operatorname{Span}\left(v_1, \ldots, v_{k+1}\right)\right) \;\geq\; (n-k)+(k+1)-n \;\geq\; 1\,.$$

Let $z \in \ker(A) \cap \operatorname{Span}\left(v_1, \ldots, v_{k+1}\right) \setminus \{0\}$. Then,

$$\begin{aligned}
\|(A-B)z\|_2^2 \;=\; \|Az\|_2^2 \;=\; \langle A^*Az, z\rangle \;&\geq\; \min_{z \in \operatorname{Span}(v_1,\ldots,v_{k+1})\setminus\{0\}} \langle A^*Az, z\rangle \\
&\geq\; \left(\min_{z' \in \operatorname{Span}(v_1,\ldots,v_{k+1})\setminus\{0\}} \mathcal{R}_{A^*A}(z')\right) \cdot \|z\|_2^2 \\
&\geq\; \sigma_{k+1}^2 \cdot \|z\|_2^2\,.
\end{aligned}$$

Thus, there exists a $z \neq 0$ such that $\|(A-B)z\|_2 \geq \sigma_{k+1} \cdot \|z\|_2$, which implies $\|A-B\|_2 \geq \sigma_{k+1}$. ■